UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/791,897	03/04/2004	Takahiko Kawatani	1509-477	4562

22879   7590   04/15/2008
HEWLETT PACKARD COMPANY
P O BOX 272400, 3404 E. HARMONY ROAD
INTELLECTUAL PROPERTY ADMINISTRATION
FORT COLLINS, CO 80527-2400

EXAMINER
PHAM, MICHAEL

ART UNIT	PAPER NUMBER
2167	

NOTIFICATION DATE	DELIVERY MODE
04/15/2008	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

JERRY.SHORMA@HP.COM
mkraft@hp.com
ipa.mail@hp.com

PTOL-90A (Rev. 04/07)

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.
- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
  Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *16 January 2008*.
2a)☐ This action is **FINAL**.   2b)☒ This action is non-final.
3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-31* is/are pending in the application.
     4a) Of the above claim(s) _____ is/are withdrawn from consideration.
5)☐ Claim(s) _____ is/are allowed.
6)☒ Claim(s) *1-8 and 13-31* is/are rejected.
7)☐ Claim(s) *9-12* is/are objected to.
8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☒ The specification is objected to by the Examiner.
10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.
     Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
     Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☒ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
     a)☐ All   b)☒ Some *   c)☐ None of:
       1.☐ Certified copies of the priority documents have been received.
       2.☐ Certified copies of the priority documents have been received in Application No. _____.
       3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
     * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO/SB/08)
     Paper No(s)/Mail Date *1/16/08*.
4) ☐ Interview Summary (PTO-413)
     Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____ .

Detailed Action

Status of claims

1. Claims 1-31 are pending.

2. Claims 1-31 have been examined.

Priority

3.   Applicant's have claimed foreign priority to JP 2003-105867 and JP 2004-30629. JP 200430629, contains a certified copy of foreign priority, and a translation. JP 2003105867, contains a certified copy for foreign priority however no translation.

Should applicant desire to obtain the benefit of foreign priority under 35 U.S.C. 119(a)-(d) prior to declaration of an interference, a certified English translation of the foreign application must be submitted in reply to this action. 37 CFR 41.154(b) and 41.202(e).

Failure to provide a certified translation may result in no benefit being accorded for the non-English application.

Specification

4.   The title of the invention is not descriptive. A new title is required that is clearly indicative of the invention to which the claims are directed.

Allowable Subject Matter

5.   The indicated allowability of claims 7 is withdrawn in view of the newly discovered

reference(s) to US Patent Application Publication 2003/0028558 by Takahiko Kawatani

(hereafter '558). Rejections based on the newly cited reference(s) follow.

6.   Claims 9-12 are objected to as being dependent upon a rejected base claim, but would be

allowable if

        a.   rewritten in independent form including all of the limitations of the base claim

and any intervening claims; and

        b.   rewritten or amended to overcome the rejection(s) under 35 U.S.C. 112, 2nd

paragraph, set forth in this Office action.

Claim Rejections - 35 USC § 112

7.   The following is a quotation of the second paragraph of 35 U.S.C. 112:

        The specification shall conclude with one or more claims particularly pointing out and
        distinctly claiming the subject matter which the applicant regards as his invention.

        Claims 1-31 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for

failing to particularly point out and distinctly claim the subject matter which applicant regards as

the invention.

8.   Claim 1-31, in particular claims 1, 7, and 29 recite in step c which provides for the use of

"using information based on the document or pattern frequency matrix for the input document or

pattern set, information based on the document or pattern frequency matrix for the input

document or pattern set, information based on the document or pattern frequency matrix for

documents or patterns in the current cluster and information based on a common co-occurrence matrix of the current cluster", but, since the claim does not set forth any steps involved in the method/process, it is unclear what method/process applicant is intending to encompass. A claim is indefinite where it merely recites a use without any active, positive steps delimiting how this use is actually practiced. All other claims fail to resolve the deficiencies of the claims from which they depend.

9.   Claim 7 is rejected under 35 U.S.C. 112, second paragraph, as being incomplete for omitting essential steps, such omission amounting to a gap between the steps. See MPEP § 2172.01. The omitted steps are: It is unclear what co-occurrence matrix S^r determines, with respect to claim 7. Further, it is unclear how the co-occurrence matrix S^r links to the other limitations of claim 7. Instead, it appears that a co-occurrence matrix is defined without indicating how the co-occurrence matrix relates to the limitations of claim 7. Thus there is a use of S^r without indicating the method steps of how or where the co-occurrence matrix S^r.

10.   Claim 9 is rejected under 35 U.S.C. 112, second paragraph as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention. The S^r equation is utilized in determining the common co-occurrence matrix T^A on the basis of a matrix T. However, it is unclear what co-occurrence matrix S^r and T actually determine.

11.   Claim 9 is rejected under 35 U.S.C. 112, second paragraph as being indefinite for failing

to particularly point out and distinctly claim the subject matter which applicant regards as the

invention. While A denotes a threshold, it is unclear of what type of threshold A determines.

The claim is therefore indefinite.

12.   Claim 9 is rejected under 35 U.S.C. 112, second paragraph as being indefinite for failing

to particularly point out and distinctly claim the subject matter which applicant regards as the

invention. An mn component is undefined. The claim is therefore indefinite.

13.   Claim 10 is rejected under 35 U.S.C. 112, second paragraph, as being incomplete for

omitting essential steps, such omission amounting to a gap between the steps. See MPEP

§ 2172.01. The omitted steps are: As used in the claim a matrix Q^A is determined on the basis

of T^A however, the description of what Q^A is missing in regards to the claimed limitation. The

claim is therefore indefinite.

14.   Claim 11 is rejected under 35 U.S.C. 112, second paragraph as being indefinite for failing

to particularly point out and distinctly claim the subject matter which applicant regards as the

invention. com1 and comq are undefined. While they determine document commonality, the

claim is indefinite as to the difference between the significance of the "1" and the "q" in com.

15.   Claim 12 is rejected under 35 U.S.C. 112, second paragraph as being indefinite for failing

to particularly point out and distinctly claim the subject matter which applicant regards as the

invention. com1 and comq are undefined. While they determine document commonality, the claim is indefinite as to the difference between the significance of the "1" and the "q" in com.

Claim Rejections - 35 USC § 101

16.   35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

17.   Claim 1-31 is rejected under 35 U.S.C. 101 because the claimed recitation of a use, without setting forth any steps involved in the process, results in an improper definition of a process, i.e., results in a claim which is not a proper process claim under 35 U.S.C. 101. See for example *Ex parte Dunki*, 153 USPQ 678 (Bd.App. 1967) and *Clinical Products, Ltd.* v. *Brenner*, 255 F. Supp. 131, 149 USPQ 475 (D.D.C. 1966). In particular, for example, claims 1, 7, and 29 recite in step c, "using information based on the document or pattern frequency matrix for the input document or pattern set, information based on the document or pattern frequency matrix for the input document or pattern set, information based on the document or pattern frequency matrix for documents or patterns in the current cluster and information based on a common co-occurrence matrix of the current cluster". All other claims fail to resolve the deficiencies of the claims from which they depend.

18.   Claims 17-22 are rejected under 35 U.S.C. 101 because the claimed invention is directed

to non-statutory subject matter. Claims 17-22 fails to fall within a statutory category of

invention. It is directed to the program itself, not a process occurring as a result of executing the

program, a machine programmed to operate in accordance with the program nor a manufacture

structurally and functionally interconnected with the program in a manner which enables the

program to act as a computer component and realize its functionality. It's also clearly not

directed to a composition of matter.

Claim Rejections - 35 USC § 103

19.   The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all

obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or
> described as set forth in section 102 of this title, if the differences between the subject
> matter sought to be patented and the prior art are such that the subject matter as a whole
> would have been obvious at the time the invention was made to a person having ordinary
> skill in the art to which said subject matter pertains. Patentability shall not be negatived
> by the manner in which the invention was made.

20.   Claim 1-6, 8, and 13-29 are rejected under 35 U.S.C. 103(a) as being unpatentable

over U.S. Patent 7130848 by Oosta (hereafter Oosta) further in view of U.S. Patent

Application Publication 2005/0022106 by Kawai et. al. (hereafter Kawai) and U.S. Patent

7225184 by Carrasco et. al. (hereafter Carrasco).

<u>Claim 1:</u>

Oosta discloses the following claimed limitations:

        "(a) obtaining a document or pattern frequency matrix for the set of input documents or

patterns based on occurrence frequencies of terms appearing in each document or pattern;"[col.

10 line 57, word correlation matrix is formed. Col. 11 lines 4-5, the matrix contains a number

that represent frequency with which that word pair is found together in the all of the abstracts of

the patent data set. Accordingly, obtaining a document or pattern frequency matrix (col. 10 line

57, correlation matrix) for the set of input documents or patterns (col. 11 lines 4-5, patent set)

based on occurrence frequencies of terms appearing in each document or pattern (col. 11 lines 4-

5, frequency with which that word pair is found together) is suggested]

"(C) obtaining the document or pattern commonality to the current cluster for each

document or pattern in the input document or pattern set by using information based on the

document or pattern frequency matrix for the input document or pattern set, information based

on the document or pattern frequency matrix for documents or patterns in the current cluster and

information based on the common-co matrix of the current cluster, and making documents or

patterns having the document commonality higher than a threshold belong temporarily to the

current cluster;" [col. 12 lines 24-30, the formation of a series of first technology topics

composed of one or more words that are strongly related to each other. The collection of first

technology topics is a second word matrix. Some words could be found in several first

technology topics, and the common words define relationships between first technology topics.

Accordingly, obtaining the document or pattern commonality to the current cluster (col. 12 lines

24-30, common words define relationships between first technology topics) for each document

or pattern in the input document or pattern set by using information based on the document

pattern frequency matrix for the input document or pattern (col. 11 lines 44-46, first technology

topics be formed by associating high frequency word pairs from the first word correlation

matrix), information based on the document or pattern frequency matrix for documents or

patterns in the current cluster (col. 11 lines 44-46, high frequency word pairs from the first

correlation matrix) and information based on the common-co matrix of the current cluster (col.

12 lines 30-34, second word matrix to further associate the related technology topics. The result

is the formation of a set of second technology topics that are condensed versions of the first

technology topics), and making documents or patterns having the document commonality higher

than a threshold belong temporally to the current cluster (col. 12 lines 11-13, use of a threshold

to form first technology topics can improve the focus of the first technology topics by

illuminating stray words.) is suggested.]

"(d) repeating step (c)" [col. 12 lines 35-40, optionally further correlations can be

conducted to form third, fourth, or fifth topics. Accordingly, (d) repeating step (c) (further

correlations conducted) is suggested]

"(f) deciding, on the basis of the document or pattern commonality of each document or

pattern to each cluster, a cluster to which each document or pattern belongs and outputting said

cluster."[col. 12 lines 53-56, assignment of a patent to a technology topic has been made based

on the number of words from a technology topic that can be found in a patent abstract.

Accordingly, deciding (assignment), on the basis of the document pattern commonality of each

document or pattern to each cluster (based on number of words), a cluster to which each

documents or pattern belongs (patent to a technology topic) and outputting said cluster (col. 11

line 34, technology topics can be formed) is suggested.]

Oosta does not explicitly disclose

"(b) selecting a seed document or pattern from remaining documents or patterns that are not included in any clustering existing at that moment and constructing a current cluster of the initial state using the seed document or pattern;"

"to extract, as the seed document or pattern, the document or pattern having the highest document or pattern commonality to the remaining documents or patterns"

"until the number of documents or patterns temporarily belong to the current cluster becomes the same as that in the previous repetition"

"(e) repeating steps (b) through (d) until a given convergence condition is satisfied; and"

On the other hand, Kawai discloses lines 12-18 of paragraph 0011, a set of candidate seed documents is evaluated to select a set of seed documents as initial cluster centers based on relative similarity between the assigned normalized score vectors for each of the candidate seed documents. The remaining non-seed documents are evaluated against the cluster centers also based on relative similarity and grouped into clusters based on a best fit, subject to a minimum fit criterion. Accordingly, Kawai discloses selecting a seed document or pattern (0011, select a set of seed documents) from remaining documents or patterns that are not included in any clustering existing at that moment (candidate seed documents) and constructing a current cluster of the initial state using the seed document or pattern (the remaining non-seed documents are evaluated against the cluster centers also based on relative similarity and are grouped into clusters).

Kawai further discloses a set of candidate seed documents is evaluated to select a set of seed documents as initial cluster centers based on relative similarity between the assigned normalized score vectors for each of the candidate seed documents. Paragraph 103, only those candidate seed documents that are sufficiently distinct from all cluster centers are selected as seed documents. Accordingly, to extract (select), as a seed document or pattern, the document or pattern (seed document) having the highest document or pattern commonality (distinct from all cluster centers) to the remaining documents or patterns (candidate seed documents).

Kawai further discloses 0101 during the first phase, seed candidate documents 60 are evaluated to identify a set of seed documents 59. In 0103, stating only those candidate seeed documents that are sufficiently distinct from all cluster centers are selected as seed documents. In 0104, if the candidate seed documents being compared are not sufficiently distinct the candidate seed is grouped into a cluster 58 with the most similar cluster center 58 to which the candidate seed document was compared. Accordingly, until the number of documents or patterns temporarily belong to the current cluster (grouped into cluster 58) becomes the same as that in the previous repetition (process continues with next seed document) is suggested.

Both Oosta and Kawai are directed towards systems capable of clustering documents. They are therefore within the same field of endeavor. For the above reasons, it would have been obvious to one of an ordinary skill in the art to have applied Kawai's disclosure above to the system of Oosta for the purpose of providing potential categories for clustering quickly, by using seed documents, and improving accuracy of clustering by pruning the candidate seed documents.

The combination of Oosta and Kawai discloses

"(e) repeating steps (b) through (d) until a given convergence condition is satisfied; and"

As Oosta discloses col. 12 lines 53-56, figure 2 element 080, identify word pair groups that form

technology topics. Accordingly, repeating step (c) until a given convergence condition is

satisfied is (amount of identified word pair groups, the more topics that are formed, hence

repeats creation of topics until all identified word groups are made) suggested.

And Kawai discloses figure 14 element 169. Hence, according to Kawai repeating steps (b) and

(d) until a given convergence condition is satisfied (e.g. last candidate seed document is met).

Oosta and Kawai do not explicitly disclose "constructing a common co-occurrence matrix of the

remaining documents or patterns" and "using the common co-occurrence matrix"

On the other hand, Carrasco discloses calculating a co-occurrence matrix of terms in common,

see claim 64. Further disclosing col. 5 lines 66-67, the matrix M is a matrix of terms in common.

col. 6 lines 53-55, recalculating from a matrix M of the remaining terms in common.

Accordingly, constructing a common co-occurrence matrix (matrix of terms in common) of the

remaining documents or patterns (remaining terms in common) is disclosed.

Carrasco further discloses col. 6 lines 1-2, from the terms-in-common matrix m, a matrix C of

correlation coefficients is constructed, as shown in figure 6. Accordingly, "using the common co-

occurrence matrix" (from the terms-in-common matrix m).

Oosta, Kawaii, and Carrasco all are directed towards clustering systems, and are thus within the

same field of endeavor. it would have been obvious to a person of an ordinary skill at the time

the invention was made to have applied Carrasco's disclosure above to the combination of Oosta

and Kawaii for the purpose of further clustering of objects and improving search by utilizing the

common co-occurrence of terms.

Claim 2:

The combination of Oosta, Kawai, and Carrasco disclose:

"(a-1) generating a document or pattern segment vector for each of said document or

pattern segments based on occurrence frequencies of terms appearing in each document or

pattern segment;" [Oosta, col. 10 lines 58, word correlation matrix]

"(a-2) obtaining a co-occurrence matrix for each document or pattern in the input

document or pattern set from the document or pattern segment vectors; and"[Oosta, col. 12 lines

24-27, series of first technology topics composed of one or more words that are strongly related

to each other. The collection of first technology topics is a second word matrix]

"(a-3) obtaining a document or pattern frequency matrix from the co-occurrence matrix for each

document." [Oosta, col. 11 lines 4-5, cell of the matrix contains a number that represent the

frequency with which that word pair is found together]

Claim 3:

The combination of Oosta, Kawaii, and Carrasco disclose "wherein, in step (b), said current cluster of the initial state is constructed to include the seed document or pattern and the neighbor documents or patterns similar to the seed document or pattern."[Kawai, paragraph 103]

Claim 4:

The combination of Oosta, Kawai, Carrasco disclose:

"(c-1) constructing a common co-occurrence matrix of the current cluster and a document or pattern frequency matrix of the current cluster;" [Oosta, col. 12 line 27, second word matrix]

"(c-2) obtaining the distinctiveness of each term and each term pair to the current cluster by comparing the document or pattern frequency matrix of the input document or pattern set and the document or pattern frequency matrix of the current cluster; and"[Oosta, col. 12 lines 32-34, the result is the formation of a set of secondary technology topics that are a condensed versions of the first technology topics. Col. 12 lines 11-13, use of threshold to form first technology topics can improve the focus of the first technology topics]

"(c-3) obtaining document or pattern commonalities to the current cluster for each document or pattern in the input document or pattern set by using the common co-occurrence matrix of the current cluster and weights of each term and term pair obtained from their distinctiveness, and making a document or pattern having the document or pattern commonality higher than a threshold belong temporarily to the current cluster." [Oosta Col. 12 lines Col. 12

lines 11-13, use of threshold to form first technology topics can improve the focus of the first

technology topics]

Claim 5:

The combination of Oosta, Kawai, and Carrasco disclose:

"repeating step (e) until the number of documents or patterns whose document or pattern

commonalities to any current clusters are less than a threshold becomes 0, or the number is less

than a threshold and is equal to that of the previous repetition."[Kawai, until next document is

empty see, figure 14, element 176]

Claim 6:

The combination of Oosta, Kawai, and Carrasco disclose:

"checking existence of a redundant cluster, and removing, when the redundant cluster exists, the

redundant cluster and again deciding the cluster to which each document belongs." [Kawai,

figure 14 element 166]

Claim 8:

The combination of Oosta, Kawai, and Carrasco disclose:

"wherein each component of the document or pattern frequency matrix of a document or pattern

set D is the number of documents or patterns in which a corresponding component of the co-

occurrence matrix of each document or pattern in the document or pattern set D does not take a

vale of zero." [Oosta, col. 11 lines 58-60, composed of all of the words in one column with a non-zero count.]

Claim 13:

The combination of Oosta, Kawai, and Carrasco disclose:

"(a) obtaining a document or pattern commonality to the remaining document or pattern set for each document or pattern in the remaining document or pattern set by using the said common co-occurrence matrix of the remaining documents or patterns," [Oosta, assignment of a patent to technology topic has been made based on the number of words from a technology topic that can be found in a patent abstract]

"(b) extracting, as candidates of the seed of the current cluster, a specific number of documents or patterns whose document or pattern commonalities obtained by step (a) are large;" [Kawai, figure 14 element 161, identify candidate seed documents]

"(c) obtaining similarities of the respective candidates of the seed of the cluster to all documents or patterns in the input document or pattern set or in the remaining document or pattern set, and obtaining documents or patterns having similarities larger than a threshold as neighbor documents or patterns of the candidate; and" [Kawai, figure 14 element 168 and 167, group candidate seed documents into similar cluster]

"(d) selecting the candidate whose number of the neighbor documents or patterns is the largest among the candidates as the seed of the current cluster and making its neighbor documents or patterns the current cluster of the initial state."[Kawai, figure 14 element 161, identify candidate seed documents]

Claim 14:

The combination of Oosta, Kawai, and Carrasco disclose:

"detecting the distinctiveness of each term or object feature and each term pair with respect to the current cluster and detecting their weights,"[Kawai, 0047, scoring module generates scores for each of the concepts and terms based on frequencies, concept weights, structural weights, and corpus weights]

the distinctiveness and weight detecting steps including

"(a) obtaining a ratio of each component of a document or pattern frequency matrix obtained from the input document or pattern set to a corresponding component of a document or pattern frequency matrix obtained from the current cluster as a document or pattern frequency ratio of each term or feature or each term or feature pair;"[Kawai, 0013, a frequency of occurrences of t least one concept within a document retrieved from the document set]

"(b) selecting a specific number of terms or features or term or feature pairs having the smallest document or pattern frequency ratios among a specific number of terms or features or term or feature pairs having the highest document or pattern frequencies, and obtaining the average of the document or pattern frequency ratios of the selected terms or features or term or feature pairs as the average document or pattern frequency ratio;"[Kawai, 0011, candidate seed documents evaluated to select a set of seed documents as initial cluster centers based on relative similarity between the assigned normalized score vectors for each of the candidate seed documents.]

"(c) dividing the average document or pattern frequency ratio by the document or pattern frequency ratio of each term or feature or each term or feature pair as a measure of the distinctiveness of each term or feature or each term or feature pair;"[Kawai, 0048, normalized vector]

"and (d) determining the weight of each term or feature or each term or feature pair from a function having the distinctiveness measure as a variable."[Kawai, 0047, scoring module generates scores for each of the concepts and terms based on frequencies, concept weights, structural weights, and corpus weights]

Claim 15:

The combination of Oosta, Kawai, and Carrasco disclose:

"eliminating terms or features and term or feature pairs having document or pattern frequencies higher than a threshold."[Oosta, col. 12 lines 4-6, a threshold can be set to accept word pairs into a first technology topic only if the count for that word is above the threshold]

Claim 16:

The combination of Oosta, Kawai, and Carrasco disclose: "wherein clustering is performed recursively by letting the document or pattern set included in a cluster be the input document or pattern set." [Kawai, Figure 14 element 168, group candidate seed document into most similar cluster]

Claim 17:

The combination of Oosta, Kawai, and Carrasco discloses: "A computer program product for causing a computer to perform the method of claim 1" [Oosta col. 19 line 67, pc].

Claim 18:

The combination of Oosta, Kawai, and Carrasco discloses: "A computer program product for

causing a computer to perform the method of claim 2"[Oosta col. 19 line 67, pc].

Claim 19:

The combination of Oosta, Kawai, and Carrasco discloses: "A computer program product for

causing a computer to perform the method of claim 3"[Oosta col. 19 line 67, pc].

Claim 20:

The combination of Oosta, Kawai, and Carrasco discloses: "A computer program product for

causing a computer to perform the method of claim 4"[Oosta col. 19 line 67, pc].

Claim 21:

The combination of Oosta, Kawai, and Carrasco discloses: "A computer program product for

causing a computer to perform the method of claim 5" [Oosta col. 19 line 67, pc].

Claim 22:

The combination of Oosta, Kawai, and Carrasco discloses: "A computer program product for

causing a computer to perform the method of claim 6" [Oosta col. 19 line 67, pc].

Claim 23:

The combination of Oosta, Kawai, and Carrasco discloses: "A computer arranged to perform the

method of claim 1" [Oosta col. 19 line 67, pc].

Claim 24:

The combination of Oosta, Kawai, and Carrasco discloses: "A computer arranged to perform the

method of claim 2" [Oosta col. 19 line 67, pc].

Claim 25:

The combination of Oosta, Kawai, and Carrasco discloses: "A computer arranged to perform the method of claim 3"[Oosta col. 19 line 67, pc].

Claim 26:

The combination of Oosta, Kawai, and Carrasco discloses: "A computer arranged to perform the method of claim 4" [Oosta col. 19 line 67, pc].

Claim 27:

The combination of Oosta, Kawai, and Carrasco discloses: "A computer arranged to perform the method of claim 5" [Oosta col. 19 line 67, pc].

Claim 28:

The combination of Oosta, Kawai, and Carrasco discloses: "A computer arranged to perform the method of claim 6" [Oosta col. 19 line 67, pc].

Claim 29:

Oosta discloses the following claimed limitations:

"A first unit for obtaining a document or pattern frequency matrix for the set of input documents or patterns, based on occurrence frequencies of terms appearing in each document or pattern;" [col. 10 line 57, word correlation matrix is formed. Col. 11 lines 4-5, the matrix contains a number that represent frequency with which that word pair is found together in the all of the abstracts of the patent data set. Accordingly, obtaining a document or pattern frequency matrix (col. 10 line 57, correlation matrix) for the set of input documents or patterns (col. 11 lines 4-5, patent set) based on occurrence frequencies of terms appearing in each document or pattern (col. 11 lines 4-5, frequency with which that word pair is found together) is suggested]

"a third unit for obtaining the document or pattern commonality to the current cluster for

each document or pattern in the input document or pattern set using information based on the

document or pattern frequency matrix for the input document or pattern set, information based

on the document or pattern frequency matrix for documents or patterns in the current cluster and

information based on the common co-occurrence matrix of the current cluster and means for

making documents or patterns having the document or pattern commonality higher than a

threshold belong temporarily to the current cluster;" [col. 12 lines 24-30, the formation of a

series of first technology topics composed of one or more words that are strongly related to each

other. The collection of first technology topics is a second word matrix. Some words could be

found in several first technology topics, and the common words define relationships between

first technology topics. Accordingly, obtaining the document or pattern commonality to the

current cluster (col. 12 lines 24-30, common words define relationships between first technology

topics) for each document or pattern in the input document or pattern set by using information

based on the document pattern frequency matrix for the input document or pattern (col. 11 lines

44-46, first technology topics be formed by associating high frequency word pairs from the first

word correlation matrix), information based on the document or pattern frequency matrix for

documents or patterns in the current cluster (col. 11 lines 44-46, high frequency word pairs from

the first correlation matrix) and information based on the common-co matrix of the current

cluster (col. 12 lines 30-34, second word matrix to further associate the related technology

topics. The result is the formation of a set of second technology topics that are condensed

versions of the first technology topics), and making documents or patterns having the document

commonality higher than a threshold belong temporally to the current cluster (col. 12 lines 11-

13, use of a threshold to form first technology topics can improve the focus of the first

technology topics by illuminating stray words.) is suggested.]

"a fourth unit for repeating the operations of the third unit" [col. 12 lines 35-40,

optionally further correlations can be conducted to form third, fourth, or fifth topics.

Accordingly, (d) repeating step (c) (further correlations conducted) is suggested]

"a sixth unit for deciding, on the basis of the document or pattern commonality of each

document or pattern to each cluster, a cluster to which each document or pattern belongs, and for

outputting said cluster." [col. 12 lines 53-56, assignment of a patent to a technology topic has

been made based on the number of words from a technology topic that can be found in a patent

abstract. Accordingly, deciding (assignment), on the basis of the document pattern commonality

of each document or pattern to each cluster (based on number of words), a cluster to which each

documents or pattern belongs (patent to a technology topic) and outputting said cluster (col. 11

line 34, technology topics can be formed) is suggested.]

Oosta does not explicitly disclose,

"second unit for selecting a seed document or pattern from remaining documents or

patterns that are not included in any cluster existing at that moment and constructing a current

cluster of the initial state using the seed document or pattern;"

"to extract, as the seed document or pattern, the document or pattern having the highest

document or pattern commonality to the remaining documents or patterns;"

"until the number of documents or patterns temporarily belonging to the current cluster becomes the same as that in the previous repetition;"

"a fifth unit for repeating the operations of the second through fourth units until given convergence conditions are satisfied; and"

On the other hand, Kawai discloses lines 12-18 of paragraph 0011, a set of candidate seed documents is evaluated to select a set of seed documents as initial cluster centers based on relative similarity between the assigned normalized score vectors for each of the candidate seed documents. The remaining non-seed documents are evaluated against the cluster centers also based on relative similarity and grouped into clusters based on a best fit, subject to a minimum fit criterion. Accordingly, Kawai discloses selecting a seed document or pattern (0011, select a set of seed documents) from remaining documents or patterns that are not included in any clustering existing at that moment (candidate seed documents) and constructing a current cluster of the initial state using the seed document or pattern (the remaining non-seed documents are evaluated against the cluster centers also based on relative similarity and are grouped into clusters).

Kawai further discloses a set of candidate seed documents is evaluated to select a set of seed documents as initial cluster centers based on relative similarity between the assigned normalized score vectors for each of the candidate seed documents. Paragraph 103, only those candidate seed documents that are sufficiently distinct from all cluster centers are selected as seed documents. Accordingly, to extract (select), as a seed document or pattern, the document or

pattern (seed document) having the highest document or pattern commonality (distinct from all

cluster centers) to the remaining documents or patterns (candidate seed documents).

Kawai further discloses 0101 during the first phase, seed candidate documents 60 are evaluated

to identify a set of seed documents 59. In 0103, stating only those candidate seed documents that

are sufficiently distinct from all cluster centers are selected as seed documents. In 0104, if the

candidate seed documents being compared are not sufficiently distinct the candidate seed is

grouped into a cluster 58 with the most similar cluster center 58 to which the candidate seed

document was compared. Accordingly, until the number of documents or patterns temporarily

belong to the current cluster (grouped into cluster 58) becomes the same as that in the previous

repetition (process continues with next seed document) is suggested.

Both Oosta and Kawai are directed towards systems capable of clustering documents. They are

therefore within the same field of endeavor. For the above reasons, it would have been obvious

to one of an ordinary skill in the art to have applied Kawai's disclosure above to the system of

Oosta for the purpose of providing potential categories for clustering quickly, by using seed

documents, and improving accuracy of clustering by pruning the candidate seed documents.

The combination of Oosta and Kawai discloses

        "a fifth unit for repeating the operations of the second through fourth units until given

convergence condition is satisfied; and"

As Oosta discloses col. 12 lines 53-56, figure 2 element 080, identify word pair groups that form

technology topics. Accordingly, repeating the third unit until a given convergence condition is

satisfied is (amount of identified word pair groups, the more topics that are formed, hence

repeats creation of topics until all identified word groups are made) suggested.

And Kawai discloses figure 14 element 169. Hence, according to Kawai repeating the second

unit and fourth unit until a given convergence condition is satisfied (e.g. last candidate seed

document is met)

The combination of Oosta and Kawai do not explicitly disclose:

"wherein selecting comprises constructing a common co-occurrence matrix of the

remaining documents or patterns; and"

"using the common co-occurrence matrix"

On the other hand, Carrasco discloses calculating a co-occurrence matrix of terms in common,

see claim 64. Further disclosing col. 5 lines 66-67, the matrix M is a matrix of terms in common.

col. 6 lines 53-55, recalculating from a matrix M of the remaining terms in common.

Accordingly, constructing a common co-occurrence matrix (matrix of terms in common) of the

remaining documents or patterns (remaining terms in common) is disclosed.

Carrasco further discloses col. 6 lines 1-2, from the terms-in-common matrix m, a matrix C of

correlation coefficients is constructed, as shown in figure 6. Accordingly, "using the common co-

occurrence matrix" (from the terms-in-common matrix m).

Oosta, Kawaii, and Carrasco all are directed towards clustering systems, and are thus within the

same field of endeavor. it would have been obvious to a person of an ordinary skill at the time

the invention was made to have applied Carrasco's disclosure above to the combination of Oosta

and Kawaii for the purpose of further clustering of objects and improving search by utilizing the

common co-occurrence of terms.

Claim 30:

The combination of Oosta, Kawaii, and Carrasco further disclose wherein the common co-

occurrence matrix reflects co-occurrence frequencies at which pairs of different terms co-occur

in each document or pattern of the remaining documents or patterns [Carrasco, col. 5 lines 64-67,

The value of Mij represents the number of secondary entities that occur with both the ith primary

entity and the jth primary entity. The matrix M is a matrix of terms in common.]

Claim 31:

The combination of Oosta, Kawai, and Carrasco further disclose wherein the common co-

occurrence matrix reflects co-occurrence frequencies at which pairs of different terms co-occur

in each document or pattern of the remaining documents or patterns [Carrasco, col. 5 lines 64-67,

The value of Mij represents the number of secondary entities that occur with both the ith primary

entity and the jth primary entity. The matrix M is a matrix of terms in common.]

21.   Claim 1-6, 8, 13-29, and 30-31 are rejected under 35 U.S.C. 103(a) as being

unpatentable over U.S. Patent 7130848 by Oosta (hereafter Oosta) further in view of U.S.

Patent Application Publication 2005/0022106 by Kawai et. al. (hereafter Kawai) and U.S.

Patent Application Publication 20040064438 by Ronald N. Kostoff (hereafter Kostoff).

<u>Claim 1:</u>

Oosta discloses the following claimed limitations:

"(a) obtaining a document or pattern frequency matrix for the set of input documents or

patterns based on occurrence frequencies of terms appearing in each document or pattern;"[col.

10 line 57, word correlation matrix is formed. Col. 11 lines 4-5, the matrix contains a number

that represent frequency with which that word pair is found together in the all of the abstracts of

the patent data set. Accordingly, obtaining a document or pattern frequency matrix (col. 10 line

57, correlation matrix) for the set of input documents or patterns (col. 11 lines 4-5, patent set)

based on occurrence frequencies of terms appearing in each document or pattern (col. 11 lines 4-

5, frequency with which that word pair is found together) is suggested]

"(C) obtaining the document or pattern commonality to the current cluster for each

document or pattern in the input document or pattern set by using information based on the

document or pattern frequency matrix for the input document or pattern set, information based

on the document or pattern frequency matrix for documents or patterns in the current cluster and

information based on the common-co matrix of the current cluster, and making documents or

patterns having the document commonality higher than a threshold belong temporarily to the current cluster;" [col. 12 lines 24-30, the formation of a series of first technology topics composed of one or more words that are strongly related to each other. The collection of first technology topics is a second word matrix. Some words could be found in several first technology topics, and the common words define relationships between first technology topics. Accordingly, obtaining the document or pattern commonality to the current cluster (col. 12 lines 24-30, common words define relationships between first technology topics) for each document or pattern in the input document or pattern set by using information based on the document pattern frequency matrix for the input document or pattern (col. 11 lines 44-46, first technology topics be formed by associating high frequency word pairs from the first word correlation matrix), information based on the document or pattern frequency matrix for documents or patterns in the current cluster (col. 11 lines 44-46, high frequency word pairs from the first correlation matrix) and information based on the common-co matrix of the current cluster (col. 12 lines 30-34, second word matrix to further associate the related technology topics. The result is the formation of a set of second technology topics that are condensed versions of the first technology topics), and making documents or patterns having the document commonality higher than a threshold belong temporally to the current cluster (col. 12 lines 11-13, use of a threshold to form first technology topics can improve the focus of the first technology topics by illuminating stray words.) is suggested.]

"(d) repeating step (c)" [col. 12 lines 35-40, optionally further correlations can be conducted to form third, fourth, or fifth topics. Accordingly, (d) repeating step (c) (further correlations conducted) is suggested]

"(f) deciding, on the basis of the document or pattern commonality of each document or

pattern to each cluster, a cluster to which each document or pattern belongs and outputting said

cluster."[col. 12 lines 53-56, assignment of a patent to a technology topic has been made based

on the number of words from a technology topic that can be found in a patent abstract.

Accordingly, deciding (assignment), on the basis of the document pattern commonality of each

document or pattern to each cluster (based on number of words), a cluster to which each

documents or pattern belongs (patent to a technology topic) and outputting said cluster (col. 11

line 34, technology topics can be formed) is suggested.]

Oosta does not explicitly disclose

"(b) selecting a seed document or pattern from remaining documents or patterns that are

not included in any clustering existing at that moment and constructing a current cluster of the

initial state using the seed document or pattern;"

"to extract, as the seed document or pattern, the document or pattern having the highest

document or pattern commonality to the remaining documents or patterns"

"until the number of documents or patterns temporarily belong to the current cluster

becomes the same as that in the previous repetition"

"(e) repeating steps (b) through (d) until a given convergence condition is satisfied; and"

On the other hand, Kawai discloses lines 12-18 of paragraph 0011, a set of candidate seed

documents is evaluated to select a set of seed documents as initial cluster centers based on

relative similarity between the assigned normalized score vectors for each of the candidate seed documents. The remaining non-seed documents are evaluated against the cluster centers also based on relative similarity and grouped into clusters based on a best fit, subject to a minimum fit criterion. Accordingly, Kawai discloses selecting a seed document or pattern (0011, select a set of seed documents) from remaining documents or patterns that are not included in any clustering existing at that moment (candidate seed documents) and constructing a current cluster of the initial state using the seed document or pattern (the remaining non-seed documents are evaluated against the cluster centers also based on relative similarity and are grouped into clusters).

On the other hand, Kawai discloses a set of candidate seed documents is evaluated to select a set of seed documents as initial cluster centers based on relative similarity between the assigned normalized score vectors for each of the candidate seed documents. Paragraph 103, only those candidate seed documents that are sufficiently distinct from all cluster centers are selected as seed documents. Accordingly, to extract (select), as a seed document or pattern, the document or pattern (seed document) having the highest document or pattern commonality (distinct from all cluster centers) to the remaining documents or patterns (candidate seed documents).

On the other hand, Kawai discloses 0101 during the first phase, seed candidate documents 60 are evaluated to identify a set of seed documents 59. In 0103, stating only those candidate seeed documents that are sufficiently distinct from all cluster centers are selected as seed documents. In 0104, if the candidate seed documents being compared are not sufficiently distinct the candidate seed is grouped into a cluster 58 with the most similar cluster center 58 to which the

candidate seed document was compared. Accordingly, until the number of documents or

patterns temporarily belong to the current cluster (grouped into cluster 58) becomes the same as

that in the previous repetition (process continues with next seed document) is suggested.

Both Oosta and Kawai are directed towards systems capable of clustering documents. They are

therefore within the same field of endeavor. For the above reasons, it would have been obvious

to one of an ordinary skill in the art to have applied Kawai's disclosure above to the system of

Oosta for the purpose of providing potential categories for clustering quickly, by using seed

documents, and improving accuracy of clustering by pruning the candidate seed documents.

The combination of Oosta and Kawai discloses

"(e) repeating steps (b) through (d) until a given convergence condition is satisfied; and"

As Oosta discloses col. 12 lines 53-56, figure 2 element 080, identify word pair groups that form

technology topics. Accordingly, repeating step (c) until a given convergence condition is

satisfied is (amount of identified word pair groups, the more topics that are formed, hence

repeats creation of topics until all identified word groups are made) suggested.

And Kawai discloses figure 14 element 169. Hence, according to Kawai repeating steps (b) and

(d) until a given convergence condition is satisfied (e.g. last candidate seed document is met).

Oosta and Kawai do not explicitly disclose "constructing a common co-occurrence matrix of the remaining documents or patterns" and "using the common co-occurrence matrix"

On the other hand, Kostoff discloses 0039 lines 1-2, a taxonomy may be developed from a collection of documents. Kostoff further discloses, 0039 line 7-8, to generate a co-occurrence matrix of high technical content phrases. The matrix cell values are then normalized and text elements are grouped, using clustering techniques, on the normalized matrix. Support can be found by the provisional application's specification on page 5 lines 9-12 and related sections.

Accordingly, Kostoff discloses a constructing (generate) a common co-occurrence matrix (co-occurrence matrix) of the remaining documents or patterns (collection of documents).

Accordingly, Kostoff further discloses using the common co-occurrence matrix (the matrix cells are then normalized and text elements grouped)

Oosta, Kawaii, and Kostoff all are directed towards clustering systems, and are thus within the same field of endeavor. It would have been obvious to a person of an ordinary skill at the time the invention was made to have applied Kostoff's disclosure above to the combination of Oosta and Kawaii for the purpose of providing phrase frequencies of occurrence within each group to indicate a level of emphasis of each group.

Claim 2:

The combination of Oosta, Kawai, and Kostoff disclose:

"(a-1) generating a document or pattern segment vector for each of said document or

pattern segments based on occurrence frequencies of terms appearing in each document or

pattern segment;" [Oosta, col. 10 lines 58, word correlation matrix]

"(a-2) obtaining a co-occurrence matrix for each document or pattern in the input

document or pattern set from the document or pattern segment vectors; and"[Oosta, col. 12 lines

24-27, series of first technology topics composed of one or more words that are strongly related

to each other. The collection of first technology topics is a second word matrix]

"(a-3) obtaining a document or pattern frequency matrix from the co-occurrence matrix for each

document." [Oosta, col. 11 lines 4-5, cell of the matrix contains a number that represent the

frequency with which that word pair is found together]

Claim 3:

The combination of Oosta, Kawaii, and Kostoff disclose "wherein, in step (b), said current

cluster of the initial state is constructed to include the seed document or pattern and the neighbor

documents or patterns similar to the seed document or pattern."[Kawai, paragraph 103]

Claim 4:

The combination of Oosta, Kawai, Kostoff discloses:

"(c-1) constructing a common co-occurrence matrix of the current cluster and a document

or pattern frequency matrix of the current cluster;" [Oosta, col. 12 line 27, second word matrix]

"(c-2) obtaining the distinctiveness of each term and each term pair to the current cluster by comparing the document or pattern frequency matrix of the input document or pattern set and the document or pattern frequency matrix of the current cluster; and"[Oosta, col. 12 lines 32-34, the result is the formation of a set of secondary technology topics that are a condensed versions of the first technology topics. Col. 12 lines 11-13, use of threshold to form first technology topics can improve the focus of the first technology topics]

"(c-3) obtaining document or pattern commonalities to the current cluster for each document or pattern in the input document or pattern set by using the common co-occurrence matrix of the current cluster and weights of each term and term pair obtained from their distinctiveness, and making a document or pattern having the document or pattern commonality higher than a threshold belong temporarily to the current cluster." [Oosta Col. 12 lines Col. 12 lines 11-13, use of threshold to form first technology topics can improve the focus of the first technology topics]

Claim 5:

The combination of Oosta, Kawai, and Kostoff discloses:

"repeating step (e) until the number of documents or patterns whose document or pattern commonalities to any current clusters are less than a threshold becomes 0, or the number is less than a threshold and is equal to that of the previous repetition."[Kawai, until next document is empty see, figure 14, element 176]

Claim 6:

The combination of Oosta, Kawai, and Kostoff discloses:

"checking existence of a redundant cluster, and removing, when the redundant cluster exists, the redundant cluster and again deciding the cluster to which each document belongs." [Kawai, figure 14 element 166]

Claim 8:

The combination of Oosta, Kawai, and Kostoff disclose:

"wherein each component of the document or pattern frequency matrix of a document or pattern set D is the number of documents or patterns in which a corresponding component of the co-occurrence matrix of each document or pattern in the document or pattern set D does not take a vale of zero." [Oosta, col. 11 lines 58-60, composed of all of the words in one column with a non-zero count.]

Claim 13:

The combination of Oosta, Kawai, and Kostoff discloses:

"(a) obtaining a document or pattern commonality to the remaining document or pattern set for each document or pattern in the remaining document or pattern set by using the said common co-occurrence matrix of the remaining documents or patterns," [Oosta, assignment of a patent to technology topic has been made based on the number of words from a technology topic that can be found in a patent abstract]

"(b) extracting, as candidates of the seed of the current cluster, a specific number of documents or patterns whose document or pattern commonalities obtained by step (a) are large;" [Kawai, figure 14 element 161, identify candidate seed documents]

"(c) obtaining similarities of the respective candidates of the seed of the cluster to all documents or patterns in the input document or pattern set or in the remaining document or pattern set, and obtaining documents or patterns having similarities larger than a threshold as neighbor documents or patterns of the candidate; and" [Kawai, figure 14 element 168 and 167, group candidate seed documents into similar cluster]

"(d) selecting the candidate whose number of the neighbor documents or patterns is the largest among the candidates as the seed of the current cluster and making its neighbor documents or patterns the current cluster of the initial state."[Kawai, figure 14 element 161, identify candidate seed documents]

Claim 14:

The combination of Oosta, Kawai, and Kostoff discloses:

"detecting the distinctiveness of each term or object feature and each term pair with respect to the current cluster and detecting their weights,"[Kawai, 0047, scoring module generates scores for each of the concepts and terms based on frequencies, concept weights, structural weights, and corpus weights]

the distinctiveness and weight detecting steps including

"(a) obtaining a ratio of each component of a document or pattern frequency matrix obtained from the input document or pattern set to a corresponding component of a document or

pattern frequency matrix obtained from the current cluster as a document or pattern frequency ratio of each term or feature or each term or feature pair;"[Kawai, 0013, a frequency of occurrences of t least one concept within a document retrieved from the document set]

"(b) selecting a specific number of terms or features or term or feature pairs having the smallest document or pattern frequency ratios among a specific number of terms or features or term or feature pairs having the highest document or pattern frequencies, and obtaining the average of the document or pattern frequency ratios of the selected terms or features or term or feature pairs as the average document or pattern frequency ratio;"[0011, candidate seed documents evaluated to select a set of seed documents as initial cluster centers based on relative similarity between the assigned normalized score vectors for each of the candidate seed documents.]

"(c) dividing the average document or pattern frequency ratio by the document or pattern frequency ratio of each term or feature or each term or feature pair as a measure of the distinctiveness of each term or feature or each term or feature pair;"[0048, normalized vector] "and (d) determining the weight of each term or feature or each term or feature pair from a function having the distinctiveness measure as a variable."[0047, scoring module generates scores for each of the concepts and terms based on frequencies, concept weights, structural weights, and corpus weights]

Claim 15:

The combination of Oosta, Kawai, and Kostoff discloses:

"eliminating terms or features and term or feature pairs having document or pattern frequencies higher than a threshold."[Oosta, col. 12 lines 4-6, a threshold can be set to accept word pairs into a first technology topic only if the count for that word is above the threshold]

Claim 16:

The combination of Oosta, Kawai, and Kostoff discloses: "wherein clustering is performed recursively by letting the document or pattern set included in a cluster be the input document or pattern set." [Figure 14 element 168, group candidate seed document into most similar cluster]

Claim 17:

The combination of Oosta, Kawai, and Kostoff discloses: "A computer program product for causing a computer to perform the method of claim 1" [Oosta col. 19 line 67, pc].

Claim 18:

The combination of Oosta, Kawai, and Kostoff discloses: "A computer program product for causing a computer to perform the method of claim 2"[Oosta col. 19 line 67, pc].

Claim 19:

The combination of Oosta, Kawai, and Kostoff discloses: "A computer program product for causing a computer to perform the method of claim 3"[Oosta col. 19 line 67, pc].

Claim 20:

The combination of Oosta, Kawai, and Kostoff discloses: "A computer program product for causing a computer to perform the method of claim 4"[Oosta col. 19 line 67, pc].

Claim 21:

The combination of Oosta, Kawai, and Kostoff discloses: "A computer program product for causing a computer to perform the method of claim 5" [Oosta col. 19 line 67, pc].

<u>Claim 22:</u>

The combination of Oosta, Kawai, and Kostoff discloses: "A computer program product for

causing a computer to perform the method of claim 6" [Oosta col. 19 line 67, pc].

<u>Claim 23:</u>

The combination of Oosta, Kawai, and Kostoff discloses: "A computer arranged to perform the

method of claim 1" [Oosta col. 19 line 67, pc].

<u>Claim 24:</u>

The combination of Oosta, Kawai, and Kostoff discloses: "A computer arranged to perform the

method of claim 2" [Oosta col. 19 line 67, pc].

<u>Claim 25:</u>

The combination of Oosta, Kawai, and Kostoff discloses: "A computer arranged to perform the

method of claim 3"[Oosta col. 19 line 67, pc].

<u>Claim 26:</u>

The combination of Oosta, Kawai, and Kostoff discloses: "A computer arranged to perform the

method of claim 4" [Oosta col. 19 line 67, pc].

<u>Claim 27:</u>

The combination of Oosta, Kawai, and Kostoff discloses: "A computer arranged to perform the

method of claim 5" [Oosta col. 19 line 67, pc].

<u>Claim 28:</u>

The combination of Oosta, Kawai, and Kostoff discloses: "A computer arranged to perform the

method of claim 6" [Oosta col. 19 line 67, pc].

Claim 29:

Oosta discloses the following claimed limitations:

"A first unit for obtaining a document or pattern frequency matrix for the set of input

documents or patterns, based on occurrence frequencies of terms appearing in each document or

pattern;" [col. 10 line 57, word correlation matrix is formed. Col. 11 lines 4-5, the matrix

contains a number that represent frequency with which that word pair is found together in the all

of the abstracts of the patent data set. Accordingly, obtaining a document or pattern frequency

matrix (col. 10 line 57, correlation matrix) for the set of input documents or patterns (col. 11

lines 4-5, patent set) based on occurrence frequencies of terms appearing in each document or

pattern (col. 11 lines 4-5, frequency with which that word pair is found together) is suggested]

"a third unit for obtaining the document or pattern commonality to the current cluster for

each document or pattern in the input document or pattern set using information based on the

document or pattern frequency matrix for the input document or pattern set, information based

on the document or pattern frequency matrix for documents or patterns in the current cluster and

information based on the common co-occurrence matrix of the current cluster and means for

making documents or patterns having the document or pattern commonality higher than a

threshold belong temporarily to the current cluster;" [col. 12 lines 24-30, the formation of a

series of first technology topics composed of one or more words that are strongly related to each

other. The collection of first technology topics is a second word matrix. Some words could be

found in several first technology topics, and the common words define relationships between

first technology topics. Accordingly, obtaining the document or pattern commonality to the

current cluster (col. 12 lines 24-30, common words define relationships between first technology

topics) for each document or pattern in the input document or pattern set by using information

based on the document pattern frequency matrix for the input document or pattern (col. 11 lines

44-46, first technology topics be formed by associating high frequency word pairs from the first

word correlation matrix), information based on the document or pattern frequency matrix for

documents or patterns in the current cluster (col. 11 lines 44-46, high frequency word pairs from

the first correlation matrix) and information based on the common-co matrix of the current

cluster (col. 12 lines 30-34, second word matrix to further associate the related technology

topics. The result is the formation of a set of second technology topics that are condensed

versions of the first technology topics), and making documents or patterns having the document

commonality higher than a threshold belong temporally to the current cluster (col. 12 lines 11-

13, use of a threshold to form first technology topics can improve the focus of the first

technology topics by illuminating stray words.) is suggested.]

"a fourth unit for repeating the operations of the third unit" [col. 12 lines 35-40,

optionally further correlations can be conducted to form third, fourth, or fifth topics.

Accordingly, (d) repeating step (c) (further correlations conducted) is suggested]

"a sixth unit for deciding, on the basis of the document or pattern commonality of each

document or pattern to each cluster, a cluster to which each document or pattern belongs, and for

outputting said cluster." [col. 12 lines 53-56, assignment of a patent to a technology topic has

been made based on the number of words from a technology topic that can be found in a patent

abstract. Accordingly, deciding (assignment), on the basis of the document pattern commonality

of each document or pattern to each cluster (based on number of words), a cluster to which each

documents or pattern belongs (patent to a technology topic) and outputting said cluster (col. 11

line 34, technology topics can be formed) is suggested.]

Oosta does not explicitly disclose,

"second unit for selecting a seed document or pattern from remaining documents or

patterns that are not included in any cluster existing at that moment and constructing a current

cluster of the initial state using the seed document or pattern;"

"to extract, as the seed document or pattern, the document or pattern having the highest

document or pattern commonality to the remaining documents or patterns;"

"until the number of documents or patterns temporarily belonging to the current cluster

becomes the same as that in the previous repetition;"

"a fifth unit for repeating the operations of the second through fourth units until given

convergence conditions are satisfied; and"

On the other hand, Kawai discloses lines 12-18 of paragraph 0011, a set of candidate seed

documents is evaluated to select a set of seed documents as initial cluster centers based on

relative similarity between the assigned normalized score vectors for each of the candidate seed

documents. The remaining non-seed documents are evaluated against the cluster centers also

based on relative similarity and grouped into clusters based on a best fit, subject to a minimum fit

criterion. Accordingly, Kawai discloses selecting a seed document or pattern (0011, select a set

of seed documents) from remaining documents or patterns that are not included in any clustering

existing at that moment (candidate seed documents) and constructing a current cluster of the

initial state using the seed document or pattern (the remaining non-seed documents are evaluated

against the cluster centers also based on relative similarity and are grouped into clusters).

Kawai further discloses a set of candidate seed documents is evaluated to select a set of seed

documents as initial cluster centers based on relative similarity between the assigned normalized

score vectors for each of the candidate seed documents. Paragraph 103, only those candidate

seed documents that are sufficiently distinct from all cluster centers are selected as seed

documents. Accordingly, to extract (select), as a seed document or pattern, the document or

pattern (seed document) having the highest document or pattern commonality (distinct from all

cluster centers) to the remaining documents or patterns (candidate seed documents).

Kawai further discloses 0101 during the first phase, seed candidate documents 60 are evaluated

to identify a set of seed documents 59. In 0103, stating only those candidate seed documents that

are sufficiently distinct from all cluster centers are selected as seed documents. In 0104, if the

candidate seed documents being compared are not sufficiently distinct the candidate seed is

grouped into a cluster 58 with the most similar cluster center 58 to which the candidate seed

document was compared. Accordingly, until the number of documents or patterns temporarily

belong to the current cluster (grouped into cluster 58) becomes the same as that in the previous

repetition (process continues with next seed document) is suggested.

Both Oosta and Kawai are directed towards systems capable of clustering documents. They are

therefore within the same field of endeavor. For the above reasons, it would have been obvious

to one of an ordinary skill in the art to have applied Kawai's disclosure above to the system of

Oosta for the purpose of providing potential categories for clustering quickly, by using seed

documents, and improving accuracy of clustering by pruning the candidate seed documents.

The combination of Oosta and Kawai discloses

    "a fifth unit for repeating the operations of the second through fourth units until given

convergence condition is satisfied; and"

As Oosta discloses col. 12 lines 53-56, figure 2 element 080, identify word pair groups that form

technology topics. Accordingly, repeating the third unit until a given convergence condition is

satisfied is (amount of identified word pair groups, the more topics that are formed, hence

repeats creation of topics until all identified word groups are made) suggested.

And Kawai discloses figure 14 element 169. Hence, according to Kawai repeating the second

unit and fourth unit until a given convergence condition is satisfied (e.g. last candidate seed

document is met)

The combination of Oosta and Kawai do not explicitly disclose:

    "wherein selecting comprises constructing a common co-occurrence matrix of the

remaining documents or patterns; and"

    "using the common co-occurrence matrix"

On the other hand, Kostoff discloses 0039 lines 1-2, a taxonomy may be developed from a collection of documents. Kostoff further discloses, 0039 line 7-8, to generate a co-occurrence matrix of high technical content phrases. The matrix cell values are then normalized and text elements are grouped, using clustering techniques, on the normalized matrix. Support can be found by the provisional application's specification on page 5 lines 9-12 and related sections.

Accordingly, Kostoff discloses wherein selecting (a taxonomy may be developed) comprises a constructing (generate) a common co-occurrence matrix (co-occurrence matrix) of the remaining documents or patterns (collection of documents).

Accordingly, Kostoff further discloses using the common co-occurrence matrix (the matrix cells are then normalized and text elements grouped)

Oosta, Kawaii, and Kostoff all are directed towards clustering systems, and are thus within the same field of endeavor. It would have been obvious to a person of an ordinary skill at the time the invention was made to have applied Kostoff's disclosure above to the combination of Oosta and Kawaii for the purpose of providing phrase frequencies of occurrence within each group to indicate a level of emphasis of each group.

Claim 30:

The combination of Oosta, Kawaii, and Kostoff further disclose wherein the common co-occurrence matrix reflects co-occurrence frequencies at which pairs of different terms co-occur

in each document or pattern of the remaining documents or patterns [Kostoff, 0038 lines 5-8,

finding text element (phrase) frequencies and text element co-occurrences in at least the relevant

documents. Kostoff, 0039 lines 11-13 text element frequencies of occurrence within each group

are summed to indicate a level of emphasis for each group.]

Claim 31:

The combination of Oosta, Kawai, and Kostoff further disclose wherein the common co-

occurrence matrix reflects co-occurrence frequencies at which pairs of different terms co-occur

in each document or pattern of the remaining documents or patterns [Kostoff, 0038 lines 5-8,

finding text element (phrase) frequencies and text element co-occurrences in at least the relevant

documents. Kostoff, 0039 lines 11-13, text element frequencies of occurrence within each group

are summed to indicate a level of emphasis for each group.]

22.   Claim 7 is rejected under 35 U.S.C. 103(a) as being unpatentable over U.S. Patent

7130848 by Oosta (hereafter Oosta) further in view of U.S. Patent Application Publication

2005/0022106 by Kawai et. al. (hereafter Kawai) and U.S. Patent Application Publication

20030028558 by Takahiko Kawatani (hereafter '558).

 Claim 7:

Oosta discloses the following claimed limitations:

  "(a) obtaining a document or pattern frequency matrix for the set of input documents or

patterns based on occurrence frequencies of terms appearing in each document or pattern;"[col.

10 line 57, word correlation matrix is formed. Col. 11 lines 4-5, the matrix contains a number

that represent frequency with which that word pair is found together in the all of the abstracts of

the patent data set. Accordingly, obtaining a document or pattern frequency matrix (col. 10 line

57, correlation matrix) for the set of input documents or patterns (col. 11 lines 4-5, patent set)

based on occurrence frequencies of terms appearing in each document or pattern (col. 11 lines 4-

5, frequency with which that word pair is found together) is suggested]

       "(C) obtaining the document or pattern commonality to the current cluster for each

document or pattern in the input document or pattern set by using information based on the

document or pattern frequency matrix for the input document or pattern set, information based

on the document or pattern frequency matrix for documents or patterns in the current cluster and

information based on the common-co matrix of the current cluster, and making documents or

patterns having the document commonality higher than a threshold belong temporarily to the

current cluster;" [col. 12 lines 24-30, the formation of a series of first technology topics

composed of one or more words that are strongly related to each other. The collection of first

technology topics is a second word matrix. Some words could be found in several first

technology topics, and the common words define relationships between first technology topics.

Accordingly, obtaining the document or pattern commonality to the current cluster (col. 12 lines

24-30, common words define relationships between first technology topics) for each document

or pattern in the input document or pattern set by using information based on the document

pattern frequency matrix for the input document or pattern (col. 11 lines 44-46, first technology

topics be formed by associating high frequency word pairs from the first word correlation

matrix), information based on the document or pattern frequency matrix for documents or

patterns in the current cluster (col. 11 lines 44-46, high frequency word pairs from the first

correlation matrix) and information based on the common-co matrix of the current cluster (col.

12 lines 30-34, second word matrix to further associate the related technology topics. The result

is the formation of a set of second technology topics that are condensed versions of the first

technology topics), and making documents or patterns having the document commonality higher

than a threshold belong temporally to the current cluster (col. 12 lines 11-13, use of a threshold

to form first technology topics can improve the focus of the first technology topics by

illuminating stray words.) is suggested.]

"(d) repeating step (c)" [col. 12 lines 35-40, optionally further correlations can be

conducted to form third, fourth, or fifth topics. Accordingly, (d) repeating step (c) (further

correlations conducted) is suggested]

"(f) deciding, on the basis of the document or pattern commonality of each document or

pattern to each cluster, a cluster to which each document or pattern belongs and outputting said

cluster."[col. 12 lines 53-56, assignment of a patent to a technology topic has been made based

on the number of words from a technology topic that can be found in a patent abstract.

Accordingly, deciding (assignment), on the basis of the document pattern commonality of each

document or pattern to each cluster (based on number of words), a cluster to which each

documents or pattern belongs (patent to a technology topic) and outputting said cluster (col. 11

line 34, technology topics can be formed) is suggested.]

Oosta does not explicitly disclose

"(b) selecting a seed document or pattern from remaining documents or patterns that are

not included in any clustering existing at that moment and constructing a current cluster of the

initial state using the seed document or pattern;"

"until the number of documents or patterns temporarily belong to the current cluster becomes the same as that in the previous repetition"

"(e) repeating steps (b) through (d) until a given convergence condition is satisfied; and"

On the other hand, Kawai discloses lines 12-18 of paragraph 0011, a set of candidate seed documents is evaluated to select a set of seed documents as initial cluster centers based on relative similarity between the assigned normalized score vectors for each of the candidate seed documents. The remaining non-seed documents are evaluated against the cluster centers also based on relative similarity and grouped into clusters based on a best fit, subject to a minimum fit criterion. Accordingly, Kawai discloses selecting a seed document or pattern (0011, select a set of seed documents) from remaining documents or patterns that are not included in any clustering existing at that moment (candidate seed documents) and constructing a current cluster of the initial state using the seed document or pattern (the remaining non-seed documents are evaluated against the cluster centers also based on relative similarity and are grouped into clusters).

On the other hand, Kawai discloses a set of candidate seed documents is evaluated to select a set of seed documents as initial cluster centers based on relative similarity between the assigned normalized score vectors for each of the candidate seed documents. Paragraph 103, only those candidate seed documents that are sufficiently distinct from all cluster centers are selected as seed documents. Accordingly, to extract (select), as a seed document or pattern, the document or pattern (seed document) having the highest document or pattern commonality (distinct from all cluster centers) to the remaining documents or patterns (candidate seed documents).

On the other hand, Kawai discloses 0101 during the first phase, seed candidate documents 60 are

evaluated to identify a set of seed documents 59. In 0103, stating only those candidate seeed

documents that are sufficiently distinct from all cluster centers are selected as seed documents.

In 0104, if the candidate seed documents being compared are not sufficiently distinct the

candidate seed is grouped into a cluster 58 with the most similar cluster center 58 to which the

candidate seed document was compared. Accordingly, until the number of documents or

patterns temporarily belong to the current cluster (grouped into cluster 58) becomes the same as

that in the previous repetition (process continues with next seed document) is suggested.

Both Oosta and Kawai are directed towards systems capable of clustering documents. They are

therefore within the same field of endeavor. For the above reasons, it would have been obvious

to one of an ordinary skill in the art to have applied Kawai's disclosure above to the system of

Oosta for the purpose of providing potential categories for clustering quickly, by using seed

documents, and improving accuracy of clustering by pruning the candidate seed documents.

The combination of Oosta and Kawai discloses

"(e) repeating steps (b) through (d) until a given convergence condition is satisfied; and"

As Oosta discloses col. 12 lines 53-56, figure 2 element 080, identify word pair groups that form

technology topics. Accordingly, repeating step (c) until a given convergence condition is

satisfied is (amount of identified word pair groups, the more topics that are formed, hence

repeats creation of topics until all identified word groups are made) suggested.

And Kawai discloses figure 14 element 169. Hence, according to Kawai repeating steps (b) and

(d) until a given convergence condition is satisfied (e.g. last candidate seed document is met).

Oosta and Kawai do not explicitly disclose wherein a co-occurrence matrix S^r of the document or

pattern Dr is determined in accordance with

$$S^r = \sum_{y=1}^{Y_r} d_{ry} d_{ry}^{\ T} \quad (1)$$

Where M is the number of sorts of the occurring terms,

D $_r$ is the r^{rth} document or pattern in a document or pattern set D consisting of R documents or

patterns,

Y_r is the number of document or pattern segments in document or pattern D $_r$, and d_{ry} =

$(d_{ry1}, .., d_{rym})^T$ is the yth document or pattern segment vector of document or pattern D $_r$ and

T represents transposition of a vector.

On the other hand, '558 discloses $S = \sum_{M=1}^{M} d_M d_M^{\ T}$

Where S is the sum matrix. 0010 discloses document segment vectors having values relating to

occurrence frequencies of terms occuring in the at least one document segment as component

values. A square sum matrix is generated from the document vectors. 0015, m = the number of

document segements. 0016, an mth document segment vector is dm=(dm1, …dmn)^T and

(m=1,…, M). 0017, N number of terms. 0018, T denotes transpose of a vector. 0019, dmn

denotes a value relating to an occurrence frequency of an n-th term occurring in the document

segment.

Accordingly, wherein a co-occurrence matrix S^r of the document or pattern Dr is determined in

accordance with

$$S^r = \sum_{y=1}^{Y_r} d_{ry} d_{ry}^{\ T} \qquad S = \sum_{M=1}^{M} d_M d_M^{\ T}$$

('558, discloses in 0013,   ) where M is the

number of sorts of the occurring terms ('558 discloses in 0017, n number of terms),

D $_r$ is the r^{rth} document or pattern in a document or pattern set D consisting of R documents or

patterns ('558 discloses 0010, input document),

Y_r is the number of document or pattern segments in document or pattern D $_r$ ('558 discloses

0016, m = number of document segments), and $d_{ry} = (d_{ry1},…, d_{rym})^T$ is the yth document or

pattern segment vector of document or pattern D $_r$ ('588 discloses 0016, mth document segment

vector is dm) and T represents transposition of a vector ('588, discloses in 0018, T represents the

transpose) is suggested.

It would have been obvious to a person of ordinary skill to have applied the disclosure of '558 to

the combination of Oosta and Kawai for the purpose of taking into account important terms,

phrases, and sentences from a document that is segmented. In doing so, allows for further

control of selecting terms, phrases, and sentences related to the central concepts of a document.

Thus, improving methods of clustering documents.

Response to Arguments

23.   Applicant's arguments with respect to claims 1-29 have been considered but are moot in

view of the new ground(s) of rejection.

24.   Applicant asserts the following directed towards the Kawai reference (lettered):

     A. the cited portions of Kawai only disclose the number or frequencies at which concepts

or terms occur individually in a document. That in contrast, the claim feature calls for

consideration of co-occurrence of terms in respective documents.

     In response to applicant's argument that the references fail to show certain features of

applicant's invention, it is noted that the features upon which applicant relies (i.e., co-occurrence

at which term A and term B co-occur in a given document) are not recited in the rejected

claim(s). Although the claims are interpreted in light of the specification, limitations from the

specification are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26

USPQ2d 1057 (Fed. Cir. 1993). Secondly, the Kawai reference discloses in paragraph 0035 a

concept is a collection of terms or phrases defining a specific meaning. Accordingly, a concept

takes into consideration of co-occurrence of terms in respective documents. Applicant's

assertions are therefore unpersuasive in this regard.

Conclusion

25.   The prior art made of record listed on PTO-892 and not relied, if any, upon is considered

pertinent to applicant's disclosure.

Contact Information

26.   Any inquiry concerning this communication or earlier communications from the

examiner should be directed to Michael D. Pham whose telephone number is (571)272-3924.

The examiner can normally be reached on Monday - Friday 9am - 5:00pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's

supervisor, John R. Cottingham can be reached on 571-272-7079. The fax phone number for the

organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent

Application Information Retrieval (PAIR) system. Status information for published applications

may be obtained from either Private PAIR or Public PAIR. Status information for unpublished

applications is available through Private PAIR only. For more information about the PAIR

system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR

system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

/M. D. P./   John R. Cottingham
Michael Pham   Art Unit 2167
Art Unit 2167   Supervisor
Examiner

/John R. Cottingham/

Supervisory Patent Examiner, Art Unit 2167